Document COO 2472-07

Annual Progress Report

A Study of Mathematical Models of Mutation and Selection in Multi-locus Systems

R. C. Lewontin

Museum of Comparative Zoology
Harvard University
Cambridge, Massachusetts 02138

Department of Energy

Contract No. DE-AS02-76EV02472

October 1, 1980 - September 30, 1981

R. C. Lewontin
Principal Investigator

Mary Marquebreuck, Director
Office for Research Contracts

A. W. Crompton, Director
Museum of Comparative Zoology

DISCLAIMER

Abstract

During the past year, research has been devoted to two related studies of two-locus systems under natural selection and one on selection in hiplo-diploid organisms. The principal results are:

1) Numerical studies were made of 2 locus selection models with asymmetric fitnesses. These were created by perturbing the fitness matrices of symmetric models whose results are known analytically. A complete classification of solved models has been made and all perturbations of these have been undertaken. The result is that all models lead to three classes of equilibrium structure. All are characterized by multiple equilibria with small linkage disequilibria under loose linkage and high complementarity equilibria under tight linkage. In some cases there is gene fixation at intermediate linkage. We believe we have established a general typology of solutions.

2) It has been shown that selection may favor more recombination, contrary to the usual expectation, if multiple locus polymorphisms are maintained by a mechanism other than marginal overdominance. This may be the result of mutation-selection balance or frequency-dependent selection.

3) In a haplo-diploid system in which diploid males are lethal (as in bees and braconid wasps) the number of sex alleles that can be maintained depends both on breeding size and the number of colonies. Simulations show that the steady number is sensitive to the number of colonies but insensitive to the number of matings. Thirty-five to fifty colonies are sufficient to maintain very large numbers of sex alleles.

General Activities

During the eight months that have elapsed since the beginning of the contract year, we have carried out both computer studies and analytic studies of three problems, two of which were planned in the previous year and the third that was added: 1) multiple locus systems with general fitnesses; 2) population genetics of recombination modifiers; and 3) brood viability in haplo-diploid organisms.

As in the previous year, work under the contract was carried out by the Principal Investigator, R. C. Lewontin, and Research Associate William Marks. In addition two graduate students, S. Orzack and L. Brooks, although not paid by the contract, continued to carry out research under it.

R. C. Lewontin made two major trips that included work and consultation under the contract. The first was to Paris where he worked for several weeks with Professor Albert Jacquard of the Institut Nationale Demographique and with Professor Michel Goux of the University of Parix VI. Professor Jacquard's group is carrying out extensive work on estimating population effective number from inbreeding statistics, a question on which we have also worked during the year. Professor Goux has been working on the estimation of sexual isolation between groups and on the problem of adaptation, both topics on which our group has also carried out research.

The second major trip was to the University of California at Davis. There the Principal Investigator consulted extensively with Professor Michael Turelli on the problem of multiple allelic heterotic stability in frequency dependent and frequency independent selection models.

R. C. Lewontin also was a participant and led a session in a three-day workshop on mathematical models in demography and genetics held at the Center for Population Studies in Cambridge under the auspices of the International Union for Demographic Studies. The proceedings of the workshop will be

published.

Finally, Professor Marcus Feldman spent four days in Cambridge with us working intensively on the two-locus general fitness model (see below) that is a joint project between the Cambridge and the Stanford groups.

Research Results

1. Two-locus models with asymmetric fitnesses

We had fully expected this model to be finished at the end of the last contract year, but the problem has turned out to be much richer in its structure than we had anticipated. In collaboration with Professor Feldman, however, we expect to have completed the exploration of the models during the remaining four months of the current contract year.

The problem is to understand the equilibrium structure of populations undergoing natural selection when two selected loci are considered simultaneously. This subject has been intensively studied during the last 15 years, chiefly by our group and Dr. Feldman, but results have been obtained largely for certain classes of symmetrical fitness relations. These include a) completely additive fitnesses, in which the fitness of a genotype is the result of adding the selection coefficients at the separate loci; b) multiplicative models in which the probability of survival for the two loci separately are multiplied; and c) the general symmetric viability model whose fitness structure is

	AA	Aa	aa
BB	$1-\gamma$	$1-\beta$	$1-\alpha$
Bb	$1-\delta$	1	$1-\delta$
bb	$1-\alpha$	$1-\beta$	$1-\gamma$

The general conclusion from these models has been that for small recombination

between locus A and locus B the equilibrium gametic array will have strong linkage disequilibrium (D large) except in the additive model when there is always linkage equilibrium (D=0). There are usually two equilibrium arrays, one with an excess of coupling gametes (D>0) and one with an excess of repulsion (D<0). As recombination, r, increases, the values of D shrink, and there may be a critical value of r, above which D≠0 equilibria disappear and are replaced by a single stable D=0 equilibrium. In some cases D=0 and D≠0 exist and are stable simultaneously.

Our work on asymmetrical models has had as its purpose a generalization of these results to arbitrary fitness models, in order to test how general the conclusions from simple models might be. The method has been to begin with a symmetric model whose solution is known analytically and then to perturb the fitnesses in a systematic way to introduce asymmetries. The results are then followed by computer solutions since analytic solutions do not exist. In last year's progress report, we discussed the outcome of the perturbation of one class of models. This is the Lewontin-Kajima super-symmetric model in which $\alpha=\delta$ in the fitness array above. Various perturbations of the α's, β's and γ's were carried out one at a time or pairwise. The results were summarized last year, but briefly we can say that D=0 is no longer in equilibrium for any model and that the bifurcation structure of the equilibria is quite different than observed in purely symmetrical models.

During the current year we have pushed these results much further. We have now perturbed 1) the additive models by making small changes in one homozygote and in simple heterozygotes; 2) the super-symmetric model in which a pair of α's in the same row were perturbed in comparison with

perturbing β; and 3)general symmetric models were generated with random α, β, γ and δ and then one α, one β, one γ and one δ were perturbed by 1%. (This part of the work was carried out by Professor Feldman.)

We have now classified all analytically solved models into 6 classes according to the nature of the epistasis, and have put each of our perturbed models into one of these. Some classes have not yet been studied, but will comprise the work for the rest of the year. Nevertheless, it now appears that all the results fall into a few quite general patterns, and we hope to be able to relate these general patterns to the epistatic structure of the models or to some other simple parametric representation of the fitness matrices. This will then produce a general theory of two-locus systems. The general equilibrium structure types are as follows:

I) For loose linkage (r~1/2) there is a single equilibrium that exists and is stable. It has a small value of D, of the order of the epistasis in asymmetrical models, reducing precisely to D=0 when there is complete symmetry. As linkage is tightened this stable equilibrium persists with D growing larger until a critical value of linkage r_c. At this point two new equilibria are born. One is unstable and has a value of D close to zero for all smaller values of r, and, in fact, D grows smaller as r grows smaller. The other is stable and is roughly complementary to the original stable equilibrium. As linkage is tightened the original D≠0 equilibrium continues on with increasing D, matched by the new stable equilibrium, also with increasing absolute value of D, but of opposite sign. Thus, for tight linkage, below r_c, there are three equilibria. Two high complementarity equilibria are stable, while the third with D~0 is unstable.

II) Type II structures are like Type I except that there is an intermediate range of recombination values, smaller than r_c, for which the high complementarity equilibria do not exist, but are replaced by gene fixation equilibria that are stable. These latter then again disappear for linkage lower than r_c and are replaced by the Type I structure.

III) There are five equilibria that exist simultaneously, three of which are interior and two of which are fixation equilibria. One of the interior equilibria has D=0, but gene frequencies are not symmetrical at both loci. For tight linkage the fixation equilibria are unstable, but they become stable as linkage loosens.

We expect all asymmetrical models to fall into the three types. All that we have investigated so far, including the perturbed additive model, do so.

2. Increase of recombination by selection

L. Brooks' work with M. Feldman on modifiers of recombination was completed. Previous work on the evolution of modifiers of recombination has shown that any mutation to a modifier of recombination between two loci that decreases the recombination will be selected for. M. Feldman has derived a general theorem that shows this to be true for an arbitrary fitness matrix if fitnesses are constant. In her new work, however, Brooks has shown that modifiers for looser linkage may be selected if the pair of loci under selection are not being maintained in equilibrium by heterosis, but by a balance between selection and mutation. This is an extremely important result. Up to the present time it has been a mystery why there is any recombination at all since selection always has seemed to favor tighter

linkage. However, much of the genetic variation we see is probably maintained by mutation-selection balance (deleterious alleles kept in the population by repeated mutation). As a result recombination may have actually been favored by selection.

A paper has been published:

Feldman, M. W., F. B. Christiansen and L. D. Brooks. 1980. Evolution of recombination in a constant environment. P.N.A.S. 77:4838-4841.

3. Two-locus frequency dependent fitnesses

A start was made on the frequency dependent fitness models at two loci, partly to see the effect of recombination in such models, but also to investigate the selection of linkage modification when equilibria are maintained by frequency dependent selection. Marginal overdominance is the rule for two-locus stable equilibria with constant fitnesses. This is one of the causes of selection always favoring tighter linkage. It is possible to create, with frequency dependent selection, stable equilibria with marginal underdominance. Such a model has been produced. For D=0 the 9 real equilibria have been solved, one of which is fully polymorphic. In this model there are also 16 complex equilibria which are of no biological interest. We will now look for $D \neq 0$ equilibria, or alter the model to create them so that the conditions for the selection of positive recombination modifiers will be created.

4. Brood viability in haploid diploids

Sex in honey bees is determined at a single locus with multiple alleles. Heterozgyotes at this locus are females. Homozygotes are diploid males;

normal males are haploid. Workers in the colony detect newly eclosed diploid males and consume them. Thus heterozygosity at this locus is lethal, and this lethality is often reflected as a severe loss of colony viability. This inbreeding load has been the single greatest deterrent to successful selective breeding programs. Working with R. E. Page, Jr. of the University of Wisconsin, W. Marks has been examining the population genetics of this system both analytically and numerically.

Assume that there are k alleles in the population, each at equal frequency (the equilibrium frequencies assuming each heterozygote to be equally fit), and that each female mates n times. Assuming random mating, it can easily be shown that the mean and variance of brood viability, V, are

$$E(V) = 1 - \frac{1}{k}$$

$$Var(V) = \frac{1}{2n}\left(\frac{2}{k}\right)\left(1 - \frac{2}{k}\right).$$

The expected brood viability depends only on the number of alleles; the variance is also inversely proportional to the number of matings.

We know further that the number of alleles which can be maintained in a finite population depends on the effective size (N_e) of the breeding population. In fact,

$$N_e = \frac{9nN_f}{4n+2}$$

in which N_f is the number of breeding females (colonies). Thus, N_e increases linearly with N_f, but quickly approaches a limit of $9N_f/4$ as the number of matings increases. Both N_f (directly) and n (by artificial insemination) can be controlled by the breeder. Clearly, it will be desirable to maintain a

large number of colonies (N_f) to maintain the maximum number of alleles over the long run. Large n (much above 10) pays a more limited return -- while it will decrease colony to colony variance in V, in the long run it will have little effect on the number of alleles maintained.

The rate at which sex alleles are lost in a finite population is also of interest. The equilibrium values may be of no particular consequence to the breeder if the rate of approach to equilibrium is very slow. To investigate this problem we have done Monte Carlo simulations with this system of sex determination, random multiple mating, and finite population size. We have shown that the rate of approach to equilibrium is very slow, depends primarily on N_f, and is relatively insensitive to n. Over the expected maximum lifetime of any breeding program (say, 40 years), N_f of 35-50 will maintain a large number of alleles, and therefore a high average V, even though at equilibrium there would be a high load.

We are in the process of examining the consequences of these sorts of non-random mating used by breeders.

5. F-statistics as estimators of population size

Effective population size can be estimated by the variance of gene frequency from generation to generation within a population. An estimate of N would be

$$\hat{N} = \frac{(\Delta p)^2}{2pq} = 2\hat{F}$$

where Δp is the change of gene frequency in one generation and p is the gene frequency before the change. Unfortunately, the observed Δp and the observed p are themselves subject to binomial sampling error which may be substantial

Recently Pamilo and Varvio-Aho and Nei have shown that a correction factor for binomial sampling error usually used gives very poor results, but unfortunately they compared the result with a model of sampling without replacement, giving rise to a hypergeometric sampling model. We have done extensive simulations of this problem assuring both sampling with and without replacement. Our finding is that no matter which correction factor is used, there is a reasonably high probability that the sampling error correction will be larger, for a given data set, than the observed gene frequency change, so that a negative or infinite estimate of N will result. None of the suggestions of new correction factors solve this problem, so estimates of N from F statistics obtained temporally remain unsatisfactory.